D754–D760 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 1 December 2011*
doi:10.1093/nar/gkr1126

Rhea—a manually curated resource of biochemical reactions

Rafael Alcántara[1,*], Kristian B. Axelsen[2], Anne Morgat[2,3], Eugeni Belda[4], Elisabeth Coudert[2], Alan Bridge[2], Hong Cao[1], Paula de Matos[1], Marcus Ennis[1], Steve Turner[1], Gareth Owen[1], Lydie Bougueleret[2], Ioannis Xenarios[2,5] and Christoph Steinbeck[1]

[1]Chemoinformatics and Metabolism Team, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK, [2]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, CMU, 1 rue Michel-Servet, CH-1211 Geneva 4, Switzerland, [3]Equipe BAMBOO, INRIA Grenoble Rhône-Alpes, 655 avenue de l'Europe, F-38330 Montbonnot Saint-Martin, France, [4]Genoscope—LABGeM, CEA, 2 Rue Gaston Crémieux, CP 5706, F-91057 Evry, France and [5]Vital-IT, Swiss Institute of Bioinformatics, Quartier Sorge, Bâtiment Génopode, CH-1015 Lausanne, Switzerland

Received August 18, 2011; Revised November 4, 2011; Accepted November 8, 2011

ABSTRACT

Rhea (http://www.ebi.ac.uk/rhea) is a comprehensive resource of expert-curated biochemical reactions. Rhea provides a non-redundant set of chemical transformations for use in a broad spectrum of applications, including metabolic network reconstruction and pathway inference. Rhea includes enzyme-catalyzed reactions (covering the IUBMB Enzyme Nomenclature list), transport reactions and spontaneously occurring reactions. Rhea reactions are described using chemical species from the Chemical Entities of Biological Interest ontology (ChEBI) and are stoichiometrically balanced for mass and charge. They are extensively manually curated with links to source literature and other public resources on metabolism including enzyme and pathway databases. This cross-referencing facilitates the mapping and reconciliation of common reactions and compounds between distinct resources, which is a common first step in the reconstruction of genome scale metabolic networks and models.

RHEA AIMS AND SCOPE

Rhea is a freely available and comprehensive resource of expert-curated biochemical reactions. It has been designed to provide a non-redundant set of chemical transformations for applications such as the functional annotation of enzymes, pathway inference and metabolic network reconstruction. Rhea provides explicit representations of biochemical reactions using chemical species from the Chemical Entities of Biological Interest ontology (ChEBI) (1) that include structural information and curated links to a host of other resources. Rhea reaction descriptions cover the official list of enzyme catalyzed reactions defined by the Nomenclature Committee of the IUBMB (NC-IUBMB) (http://www.chem.qmul.ac.uk/iubmb/enzyme/) (2,3). This extends where possible to the provision of explicit descriptions for specific instances of generic reactions that are catalyzed by enzymes with broad substrate specificity (as well as reactions that are only referred to within the free text comments of IUBMB entries). One example of such a generic reaction is that catalyzed by alcohol dehydrogenase (EC 1.1.1.1), which is described in the following way within the IUBMB classification:

$$\text{An alcohol} + \text{NAD}^+ = \text{an aldehyde or ketone} + \text{NADH}$$

Within Rhea, a generic reaction description is provided that corresponds to this textual description, along with specific reactions for all known substrate/product pairs. A similar approach is used to describe reactions involving polymers with varying numbers of repeated units, such as isoprenoid quinones, which participate in the reaction catalyzed by malate dehydrogenase (EC 1.1.5.4):

$$(S)\text{-malate} + \text{a menaquinone} = \text{oxaloacetate} + \text{a menaquinol}$$

This reaction involves a menaquinone with a variable number of isoprene units (4): eight isoprene units in

*To whom correspondence should be addressed. Tel: +44 0 1223 494414; Fax: +44 0 1223 494468; Email: rafael.alcantara@ebi.ac.uk

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

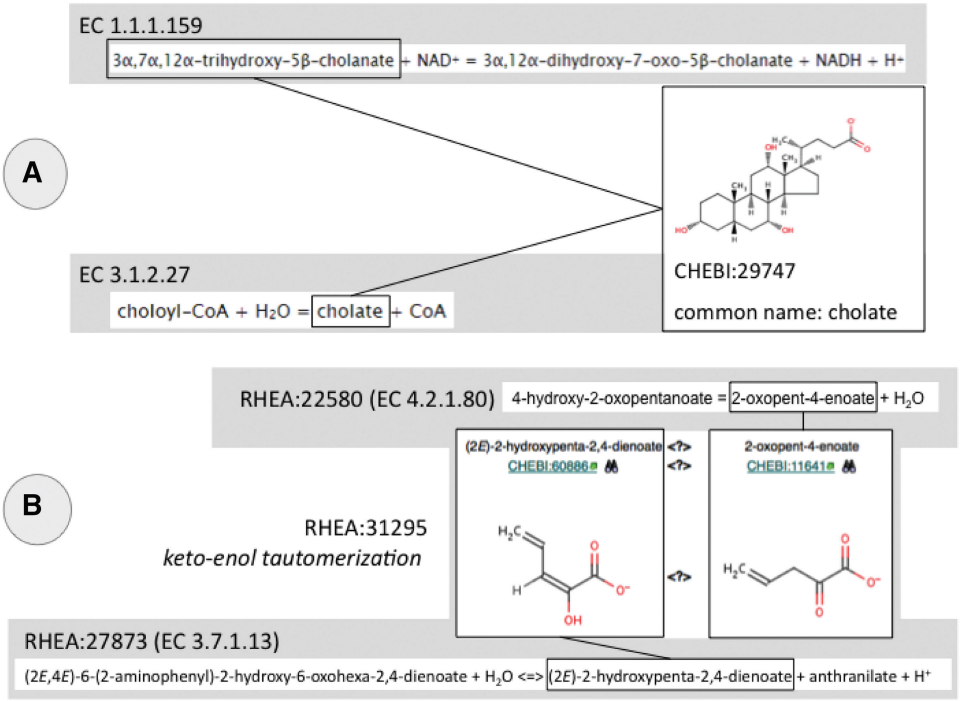

Figure 1. Chemical compound issues. (**A**) The same chemical compound can be described using different names in textual representations of reactions in the IUBMB classification. (**B**) A chemical compound may exist in different forms that can interchange spontaneously, such as keto and enol tautomers, and reactions may include each of these forms. To illustrate this, RHEA:22580 describes a reaction involving a keto tautomeric form while RHEA:27873 describes a separate reaction involving an enolic form of the same compound. The keto-enol tautomerization reaction RHEA:31295 allows the two reactions to be linked if necessary.

Escherichia coli (menaquinone-8, MK-8), seven isoprene units in certain species of the genus Bacillus and nine in species of *Streptococcus*. Each of these reactions has a precise description within Rhea, which specifies the number of repeat units within the menaquinone.

In addition to providing specific descriptions of known instances of generic reactions, Rhea also provides transport reactions and spontaneous reactions that lack a corresponding textual definition in the IUBMB classification. These reactions are designed to facilitate the use of Rhea as a reference resource for genome-scale metabolic network reconstruction, where precise definitions of substrate and product specificity, and the inclusion of spontaneous and transport reactions, are essential steps in building a functional genome-scale model for metabolism (5–7). More generally, the use of explicit reaction descriptions facilitates the precise identification, mapping and comparison of compounds and reactions between different resources and different metabolic models (8). This can be difficult to achieve using the systematic unambiguous chemical nomenclature of IUPAC (International Union of Pure and Applied Chemistry), as biologists often prefer common names to IUPAC standardized labels (which can be relatively complex). The inconsistent use of common names and standardized nomenclature can lead to ambiguity in reaction and compound descriptions that requires manual curation to resolve, as illustrated in Figure 1A, where a single compound is referred to variously as cholate and 3α,12α-dihydroxy-7-oxo-5β-cholanate within two reaction descriptions. A further source of ambiguity arises from the use of generic compound labels that are intended to describe more than one chemical species, such as NAD(P)/NAD(P)H, which is often used in oxidoreduction reactions that use NAD^+/NADH or $NADP^+$/NADPH redox couples. Rhea provides an explicit description of each of the corresponding reactions, allowing unambiguous assignment of reactions including NAD+/NADH or NADP+/NADPH. A third example of ambiguity may occur when considering reactions involving keto-enol tautomers. Rhea provides spontaneous tautomerization reactions that can be used to link reactions involving tautomeric forms of the same compound, such as RHEA:31295 which connects reactions involving (2*E*)-2-hydroxypenta-2,4-dienoate (enol form) and 2-oxopent-4-enoate (keto form) (Figure 1B).

Although Rhea attempts to reduce or eliminate ambiguity in reaction descriptions wherever possible, some allowance is made for incomplete knowledge of reaction chemistry. Rhea provides incomplete reactions where not all the reactants are known, and where the reactions are not necessarily balanced. These reactions are clearly identified by their 'preliminary' status.

Reaction representation in Rhea

Rhea provides explicit representations of biochemical reactions using chemical species from the ChEBI ontology (1). ChEBI includes information on chemical formula and charge, as well as nomenclature and 2D-structural information in various chemical formats. The latter information is used by Rhea for reaction search, display,

Table 1. The master reaction RHEA:15133 (shown in Figure 3) has an undefined direction, represented with the symbol <?>. It is associated with three directional reactions (RHEA:15134, RHEA:15135 and RHEA:15136), each of which has a specific set of corresponding cross-references to external databases

Reaction and direction	Cross-references
RHEA:15133 (master reaction) L-glutamate + H_2O + NAD^+ <?> 2-oxoglutarate + H^+ + NADH + NH_4^+	*None*
RHEA:15134 - Left-to-Right (RHEA:15133, LR) L-glutamate + H_2O + NAD^+ → 2-oxoglutarate + H^+ + NADH + NH_4^+	UniPathway:UER00591 Reactome:REACT_710.4
RHEA:15135 - Left-to-Right (RHEA:15133, LR) 2-oxoglutarate + H^+ + NADH + NH_4^+ → L-glutamate + H_2O + NAD^+	Reactome:REACT_1896.4
RHEA:15136 - Bidirectional (RHEA:15133, BI) L-glutamate + H_2O + NAD^+ ⇔ 2-oxoglutarate + H^+ + NADH + NH_4^+	UniPathway:UCR00243 KEGG:R00243 MetaCyc: GLUTAMATE-DEHYDROGENASE-RXN IntEnz: EC 1.4.1.2 IntEnz:EC 1.4.1.3 UniProt: DHE2_PORG3 UniProt: DHEA_NICPL,...

and export, as well as validation and balancing. Rhea reactions (like their constituent ChEBI entities) are manually curated and linked to a host of other resources, including underlying structural information.

Each possible reaction in Rhea is represented by a unique 'master reaction' that is independent of any biological context, having no associated directional information. Each such master reaction has the following attributes:

- A unique identifier
- Two reaction parts (left and right)
- A set of qualifiers that describe the type of reaction and if it is balanced
- A curation status (approved, preliminary or obsolete)

Each of the two reaction parts (arbitrarily defined as left and right) are composed of a set of participant *compounds*, their stoichiometric *coefficient* and possibly their *localization*. The *compounds* are defined by a ChEBI identifier, a name, a chemical formula, a net charge and possibly a 2D structure. The *coefficient* may be an integer or a symbolic expression (n, $n+1$, $n-1$, etc). For transport reactions, nominal cellular compartments (*localization*) are specified by the tokens (*in*) and (*out*) using the side convention of NC-IUBMB.

Each master reaction is uniquely represented at an arbitrarily chosen pH of 7.3. Rhea uses the Marvin pK_a calculator from ChemAxon (http://www.chemaxon.com) to select the major species of each compound found at pH 7.3 and at a temperature of 298 K (9). For some compounds, such as phosphate, the major species at pH 7.3 (which is HPO_4^{2-}) may be present in only a slight excess over other minor species (such as $H_2PO_4^-$), so the reaction representation is a simplification of the actual state. However, the selection of the major species serves to guarantee data consistency, since the same protonation state will be used in all reactions (and the reactions are fully balanced, including protons). This convention is in accord with that adopted by some existing resources, such as MetaCyc (10), but contrasts with that adopted by others, such as KEGG (11) and BRENDA (12), which generally choose a neutral form for compound representation. This latter convention has the advantage of being simple but prevents the balancing of reactions for charge and hydrogen atoms. Note that the export formats provided by Rhea allow users to compute structures at other pH values using commonly available chemoinformatics tools.

To ensure data consistency, the following additional constraints are supported by the Rhea infrastructure:

- The same compound (ChEBI ID, localization) cannot occur on both sides of a reaction. This results in the exclusion of certain common species, such as Mg^{2+} in Mg^{2+}-ATP.
- The reaction must be chemically balanced for mass and charge, i.e. the compounds found in the left and right parts of the equation must have the same total number of atoms of each type and the same net charge.
- The reaction must be unique. To ensure this uniqueness, a fingerprint is computed based on the compounds on both reaction parts (ChEBI ID, coefficient, localization).

Each unique master reaction is associated with three directional reactions—forward, reverse and bidirectional (reversible)—each of which has a distinct identifier (Table 1). (Note that the validation steps described above for the master reaction are automatically applied to each of these directional reactions.) Rhea is therefore redundant, in the sense that each reaction has four distinct (validated) representations, but this feature allows directional reactions from external resources to be linked to the appropriate directional reaction in Rhea (Table 1). The drawback of this approach is that Rhea may include directional reactions that are unfeasible in biological systems according to current knowledge.

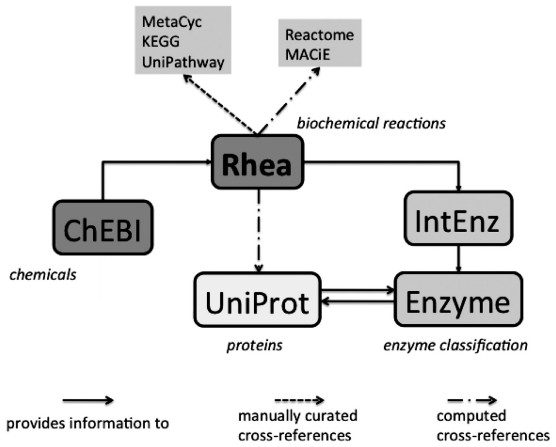

Figure 2. The relationships between Rhea and databases of chemical compounds, proteins and enzyme classification. ChEBI provides information on chemical compounds for the creation of chemical reactions in Rhea. The resulting Rhea reactions are used in IntEnz to describe the enzymes catalyzing the reactions. ENZYME is generated from IntEnz data and currently uses the textual representation reactions also used by the IUBMB EC list. Also shown are the cross-references provided by Rhea to other resources. These are generated manually and computationally.

Rhea curation

Each master reaction has a curation status, which is one of approved, preliminary or obsolete. To be approved, a reaction has to fulfill the constraints described in the previous section relating to unicity and balance. Reactions can also have descriptive qualifiers such as 'chemically balanced', 'transport', 'class of reaction' and 'polymerization'.

The manual curation process includes verification of the selected ChEBI entities (assisted by chemoinformatic tools), the addition of original literature citations, and the addition of cross references to other resources. Rhea citations are managed in the EBI's CiteXplore bibliography database (http://www.ebi.ac.uk/citexplore/). Rhea cross-references several resources describing biochemical reactions including KEGG (11), EcoCyc/MetaCyc (10,13) and UniPathway (14). As described above, these cross-references link specific directional instances of reactions in Rhea to those of other resources. Cross-references are also added automatically to Reactome (15) and MACiE (16) on the basis of their reaction participants (ChEBI compound) and the indicated reaction direction.

An important goal of the Rhea project is to link the chemical information from ChEBI to that in resources describing enzymatic reactions, such as IntEnz (Integrated relational Enzyme database) (17) that contains data on enzymes organized by EC numbers (Figure 2). Rhea now provides the chemical representation for enzyme-catalyzed reactions in IntEnz (from which Rhea is cross-referenced). Rhea also includes cross-references to EC numbers, thus providing an entry point to IntEnz and enzyme classification. EC numbers are subsequently used to propose cross-references to protein sequences in the UniProtKB/Swiss-Prot knowledgebase (18).

Rhea curation may also include the selection of descriptive names or labels for the participating ChEBI compounds, in order to provide a more easily understandable (human readable) reaction description. It is important to note that ChEBI and Rhea labels for a specific compound may differ. Rhea makes use of labels marked 'UniProt synonym' in ChEBI (these labels are so named as they will in future provide a controlled vocabulary for chemical compounds in UniProtKB). To illustrate this, CHEBI:15378 has the ChEBI common name *hydron* whereas Rhea uses the label H^+. This practice of selecting easily understood chemical labels for Rhea means that the displayed label for certain chemical species will not necessarily show the correct charge state (a typical example being NAD^+, CHEBI:57540, which is actually negatively charged). However, the underlying compounds are correctly charged and the reaction is appropriately balanced.

Comparison of Rhea and related resources

When the Rhea project was initiated, the only freely available database of reactions was the KEGG LIGAND database. During the intervening period, several additional resources containing information on chemical compounds and reactions have become available (while KEGG data now requires a subscription for download) (Table 2). Some of these newly available resources focus on enzymes catalyzing chemical reactions [e.g. BRENDA (12), ExplorEnz (2), Enzyme (3) and the aforementioned IntEnz (15)], while others provide detailed information on reaction mechanisms [e.g. EzCatDB (19), MACiE (16)] or kinetic data [e.g. BRENDA (12), SABIO-RK (20)]. Resources that provide comprehensive reaction data and are comparable in scope with Rhea include BioPath (21), KEGG (11) and MetaCyc (10).

Submission to Rhea

Rhea welcomes submissions describing reactions that are not currently available in Rhea. All new reaction submissions should be posted on our SourceForge Reaction Requests/Updates tracker (https://sourceforge.net/projects/rhea-ebi/) with relevant information including ChEBI identifiers for each reaction participant and cross-references to other relevant databases and source literature where available.

Rhea content

At the time of writing, Rhea (release 24) includes 4321 master reactions (each associated with three directional reactions) that involve 3788 distinct ChEBI chemical entities. Among them, 251 are transport reactions. In the corresponding IntEnz database (release 71) 3145 of the 4596 enzyme entries (EC numbers) have their reactions described in Rhea. This corresponds to a total of 3658 distinct Rhea reactions (as there may be a many-to-many relationship between reactions and enzymes).

The database is updated by monthly releases. Updates are synchronized with ChEBI releases.

Table 2. Resources relating to biochemical reactions

Resource	Location
BioCyc: MetaCyc (10) and EcoCyc (13)	MetaCyc (10) http://metacyc.org; EcoCyc (13) http://ecocyc.org
Biopath (21)	http://www.molecular-networks.com/biopath/
BKM-react (26)	http://bkm.tu-bs.de/
BRENDA (12)	www.brenda-enzymes.org
ENZYME (3)	http://enzyme.expasy.org/
ExplorEnz (2)	http://www.enzyme-database.org
EzCatDB (19)	http://mbs.cbrc.jp/EzCatDB/
IntEnz (17)	http://www.ebi.ac.uk/intenz
KEGG (11)	http://www.genome.jp/kegg/
MACiE (16)	http://www.ebi.ac.uk/thornton-srv/databases/MACiE/
Reactome (15)	http://www.reactome.org/
SABIO-RK (20)	http://sabio.villa-bosch.de/
UniPathway (14)	http://www.unipathway.org/pathway

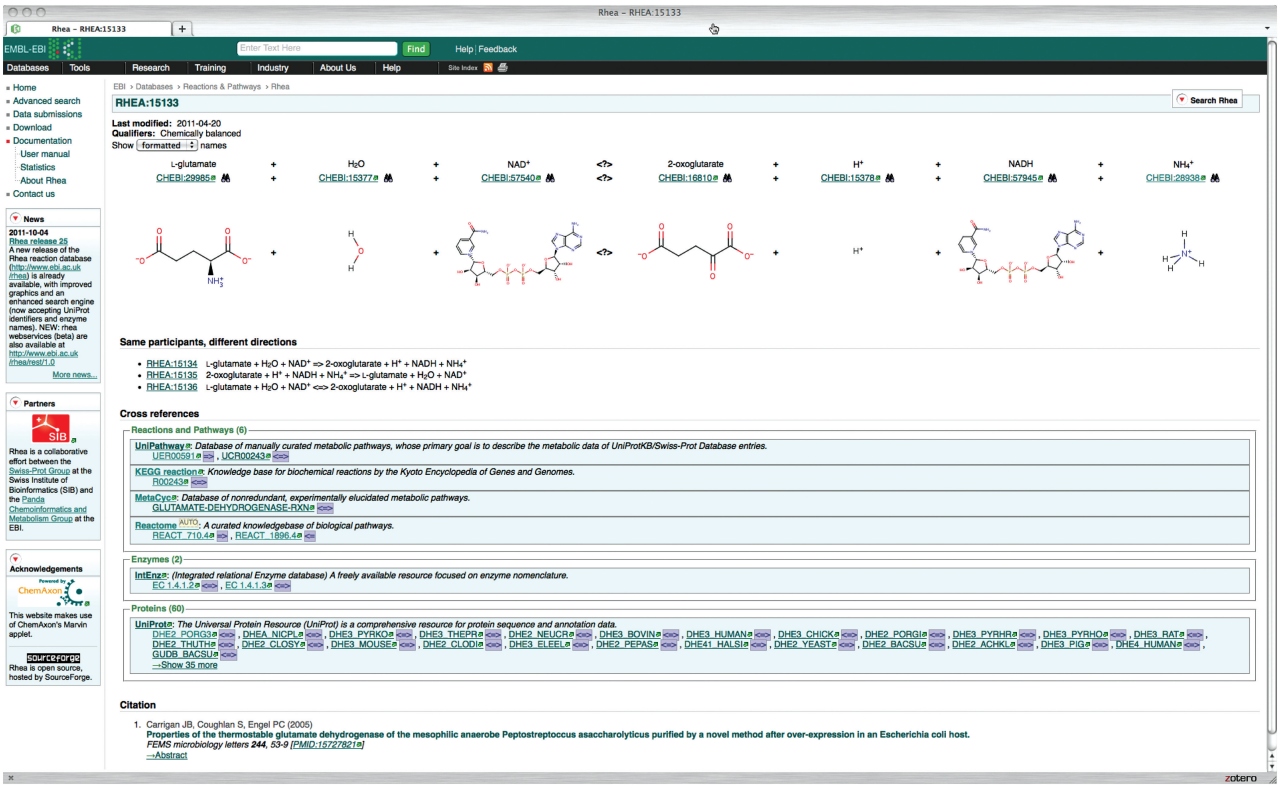

Figure 3. Sample master reaction and associated directional reactions in Rhea. http://www.ebi.ac.uk/rhea/reaction.xhtml?id=15133 RHEA:15133 is a master reaction. This master reaction is described by its chemicals (labels and identifiers in ChEBI), and has no specific direction. The three associated directional reactions are indicated in the section titled 'Same participants, different directions'. These are: RHEA:15134 (left-to-right), RHEA:15135 (right-to-left) and RHEA:15136 (bidirectional). The section titled 'Cross-references' lists all related resources to which one of the directional reactions has been linked. The actual Rhea reaction can be determined from the directional icon ('=>' for RHEA:15134, '<=' for RHEA:15135 and '<=>' for RHEA:15136). The user can retrieve a list of all the reactions a specific compound is involved in by clicking the binocular symbol next to the compound name. Marvin is used for displaying chemical structures, Marvin 5.0.0, 2011 (http://www.chemaxon.com).

Rhea web server

The Rhea web server (http://www.ebi.ac.uk/rhea) enables access to Rhea data. It provides browsing, searching, web services and download facilities.

An example of a reaction page is shown in Figure 3. It presents the chemical transformation (including the reaction participants' names, ChEBI identifiers and chemical structures), and related information (cross-references and citations). The Rhea web site provides a number of search facilities (described below), and also allows users to navigate the reaction set via their common compounds.

The Rhea website allows simple queries using any of the following as input:

- a reaction identifier from Rhea or any of the cross-referenced resources [KEGG (11), EcoCyc (13),

MetaCyc (10), UniPathway (14), MACiE (16) or Reactome (15)];
- a compound name or identifier from ChEBI. The ChEBI index is used to resolve any synonyms and cross-references;
- an equation describing the reaction. The Rhea search tool will attempt to parse this and return any matching reactions. If none is found, it will look for similar reactions, i.e. with as many of the participants in the equation as possible;
- an EC number, an enzyme name or a UniProtKB/Swiss-Prot identifier. The IntEnz index is used to resolve any enzyme synonyms;
- a bibliographic citation, or any part of, such as an author name, title, abstract or publication identifier (PubMed identifier).

In addition, an advanced search page allows a search to be restricted to specific fields (e.g. reaction participants, cross-references or citations) and to perform structural searches. In this latter case, the user can import or draw a 2D-structure through the JChemPaint applet (jchempaint.sourceforge.net), following which the chemical structure search algorithm OrChem (22) will perform a substructure or similarity search on the set of compounds involved in Rhea. The complete documentation of the structural search is available on the ChEBI web server (http://www.ebi.ac.uk/chebi/userManualForward.do).

Downloads

All Rhea data is available for free download (http://www.ebi.ac.uk/rhea/download.xhtml) in three formats. These are BioPAX level 2 (23), RXN and RD (24), which facilitate the use of data by chemoinformatics software tools.

BioPAX (23) is a collaborative effort to create a data exchange format for biological pathway data. It is defined in OWL and represented in RDF/XML syntax. Rhea reactions correspond to the biochemicalReaction, transport and transportWithBiochemicalReaction BioPax classes. An example of a BioPAX export is given in the supplementary data section (Supplementary Figure S1). RXN is one of the chemical table (CT) file formats specified by Accelrys (formerly Symyx and MDL) (24). RXN represents unidirectional processes, so the Rhea export in RXN includes only this subset of Rhea reactions. RD (reaction data) is a second CT file format, consisting of a set of records, each record defining one directional reaction (in RXN format) and associated data.

The 2D structures of the subset of ChEBI compounds referenced by Rhea are also available as an SDF file, a chemical format specified by Accelrys (formerly by MDL, 24). That format includes information about the atoms, bonds, connectivity and coordinates of each of the molecules of interest.

Web services

The Rhea resource is exposed as RESTful web services. Results following standard HTTP GET requests are sent to the client (e.g. a web browser) in one of RXN (24), BioPAX Level 2 (23) or CMLReact formats (25), depending on the HTTP Accept header or the URL of the request.

The Service makes available two methods: one for general searching of reactions and another for retrieving the full entry of a reaction.

The main use case of the Rhea Web Service is a client application, which invokes the search method to get a list of reaction URLs in the format it requires, and then accesses these URLs to retrieve the full reaction entries. Another use case is a client application, which acquires a Rhea identifier, referenced in other services and then creates the URL specifying the format it requires to retrieve the full entry of the reaction. The service URL and the instructions on how to access the service can be found at http://www.ebi.ac.uk/rhea/rest/1.0.

Software

The Rhea database architecture and software tools are distributed as Open Source (at http://sourceforge.net/projects/rhea-ebi/) allowing end-users to download and install their own local database of reactions. All software is written in Java 6.

The software package includes amongst other things:

- Database schema—allowing the storage and validation of data.
- Domain model and data validator—providing a reaction model and ability to validate the reaction.
- Rhea annotation tool—provides the ability for curators to add and modify their own reactions.
- Rhea public website—allowing reaction visualization.

The Rhea database runs on Oracle 11 g (http://www.oracle.com). However, minor tweaks allow it to run on open database platforms such as MySQL (www.mysql.com). The database schema is given in the supplementary data section (Supplementary Figure S2 and Supplementary Table S1).

Future directions

Collaborative developments with the Universal Protein Resource KnowledgeBase, UniProtKB (13), are ongoing with the aim of using Rhea as a reference vocabulary for describing enzymatic reactions within UniProtKB protein sequence records. Rhea also aims at serving as a general resource of chemical reactions for the reconstruction of genome-scale metabolic networks, as in the Microme (http://www.microme.eu) and MetaNetX (http://www.metanetx.org) initiatives. We are examining a number of curated metabolic networks in order to identify missing reactions of interest and to curate these in Rhea, with the aim of enhancing the unique content of this reaction resource. We are also developing and enhancing our submission tools to allow batch submission, thereby speeding up the population of Rhea with new reactions. We will also further exploit the chemical ontology of ChEBI, where relationships between reaction participants can serve as a basis for the development of a logical reaction classification.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online. Supplementary Figures 1–2 and Table 1.

ACKNOWLEDGEMENTS

The authors thank Amos Bairoch, Kirill Degtyarenko, Henning Hermjakob and Eric Coissac for their help and encouragement in the early stages of this project, and Andrea Auchincloss and Alain Viari for critically reading the manuscript.

FUNDING

The Swiss Federal Government through the Federal Office of Education and Science; the European Molecular Biology Laboratory (core funding); European Union (SLING: Serving Life-science Information for the Next Generation [226073], Microme: A Knowledge-Based Bioinformatics Framework for Microbial Pathway Genomics [222886-2], and ERC Advanced Grant SISYPHE); French government through ANR MIRI [BLAN08-1335497]; and the MetaNetX project of the Swiss SystemsX.ch initiative. Funding for open access charge: EMBL-EBI.

Conflict of interest statement. None declared.

REFERENCES

1. de Matos,P., Alcántara,R., Dekker,A., Ennis,M., Hastings,J., Haug,K., Spiteri,I., Turner,S. and Steinbeck,C. (2010) Chemical entities of biological interest: an update. *Nucleic Acids Res.*, **38**, D249–D254.
2. McDonald,A.G., Boyce,S. and Tipton,K.F. (2009) ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.*, **37**, D593–D597.
3. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
4. Collins,M.D. and Jones,D. (1981) Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol. Rev.*, **45**, 316–354.
5. Poolman,M.G., Bonde,B.K., Gevorgyan,A., Patel,H.H. and Fell,D.A. (2006) Challenges to be faced in the reconstruction of metabolic networks from public databases. *Syst. Biol.*, **153**, 379–384.
6. Feist,A.M., Herrgård,M.J., Thiele,I., Reed,J.L. and Palsson,B.Ø. (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.*, **7**, 129–143.
7. Radrich,K., Tsuruoka,Y., Dobson,P., Gevorgyan,A., Swainston,N., Baart,G. and Schwartz,J.M. (2010) Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. *BMC Syst. Biol.*, **4**, 114.
8. Oberhardt,M.A., Puchałka,J., Martins dos Santos,V.A. and Papin,J.A. (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Comput. Biol.*, **7**, e1001116.
9. Perrin,D.D. (1981) *pKa Prediction for Organic Acids and Bases*. Springer, Berlin.
10. Caspi,R., Altman,T., Dale,J.M., Dreher,K., Fulcher,C.A., Gilham,F., Kaipa,P., Karthikeyan,A.S., Kothari,A., Krummenacker,M. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
11. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
12. Scheer,M., Grote,A., Chang,A., Schomburg,I., Munaretto,C., Rother,M., Söhngen,C., Stelzer,M., Thiele,J. and Schomburg,D. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.*, **39**, D670–D676.
13. Keseler,I.M., Collado-Vides,J., Santos-Zavaleta,A., Peralta-Gil,M., Gama-Castro,S., Muniz-Rascado,L., Bonavides-Martinez,C., Paley,S., Krummenacker,M., Altman,T. *et al.* (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.*, **39**, D583–D590.
14. Morgat,A., Coissac,E., Coudert,E., Axelsen,K.B., Keller,G., Bairoch,A., Bridge,A., Bougueleret,L., Xenarios,I. and Viari,A. (2012) UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.*, **40**, D761–D769.
15. D'Eustachio,P. (2011) Reactome knowledgebase of human biological pathways and processes. *Methods Mol. Biol.*, **694**, 49–61.
16. Holliday,G.L., Almonacid,D.E., Bartlett,G.J., O'Boyle,N.M., Torrance,J.W., Murray-Rust,P., Mitchell,J.B. and Thornton,J.M. (2007) MACiE (mechanism, annotation and classification in enzymes): novel tools for searching catalytic mechanisms. *Nucleic Acids Res.*, **35**, D515–D520.
17. Fleischmann,A., Darsow,M., Degtyarenko,K., Fleischmann,W., Boyce,S., Axelsen,K.B., Bairoch,A., Schomburg,D., Tipton,K.F. and Apweiler,R. (2004) IntEnz, the integrated relational enzyme database. *Nucleic Acids Res.*, **32**, D434–D437.
18. The UniProt Consortium. (2011) Ongoing and future developments at the universal protein resource. *Nucleic Acids Res.*, **39**, D214–D219.
19. Nagano,N. (2005) EzCatDB: the enzyme catalytic-mechanism database. *Nucleic Acids Res.*, **33**, D407–D412.
20. Rojas,I., Golebiewski,M., Kania,R., Krebs,O., Mir,S., Weidemann,A. and Wittig,U. (2007) Storing and annotating of kinetic data. *In Silico Biol.*, **7**, S37–S44.
21. Reitz,M., Sacher,O., Tarkhov,A., Trümbach,D. and Gasteiger,J. (2004) Enabling the exploration of biochemical pathways. *Org. Biomol. Chem.*, **2**, 3226–3237.
22. Rijnbeek,M.L. and Steinbeck,C. (2010) OrChem: an open source chemistry search engine for Oracle. *J. Cheminform.*, **2**, 28.
23. Demir,E., Cary,M.P., Paley,S., Fukuda,K., Lemer,C., Vastrik,I., Wu,G., D'Eustachio,P., Schaefer,C., Luciano,J. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.*, **28**, 935–942.
24. Dalby,A., Nourse,J.G., Hounshell,W.D., Gushurst,A.K.I., Grier,D.L., Leland,B.A. and Laufer,J. (1992) Description of several chemical structure file formats used by computer programs developed at molecular design limited. *J. Chem. Inf. Comput. Sci.*, **32**, 244–255.
25. Holliday,G.L., Murray-Rust,P. and Rzepa,H.S. (2006) Chemical markup, XML, and the world wide web. 6. CMLReact, an XML vocabulary for chemical reactions. *J. Chem. Inf. Model*, **46**, 145–157.
26. Lang,M., Stelzer,M. and Schomburg,D. (2011) BKM-react, an integrated biochemical reaction database. *BMC Biochem.*, **12**, 42.